Plant Pathology Bulletin 16: 225-229, 2007

以群體測試結果估算種子傳毒率的統計方法

蔣國司[1]　鄧汀欽[2, 3]　賴信宏[1]　王淑芬[1]

[1] 台中市 國立中興大學農藝學系
[2] 台中縣 行政院農業委員會農業試驗所植物病理組
[3] 聯絡作者：電子郵件：tcde@wufeng.tari.gov.tw，傳眞：04-2333-8162
接受日期：中華民國 96 年 9 月 10 日

摘　要

蔣國司、鄧汀欽、賴信宏、王淑芬. 2007. 以群體測試結果估算種子傳毒率的統計方法. 植病會刊 16: 225-229.

爲發展經濟實用的種子 (苗) 帶毒檢測技術，並以統計方法估算不同批號種子批之傳毒比率，本文針對群體測試的結果，使用R程式語言開發出更具有彈性的模式來估算種子傳毒率，以期解決之前文獻對於此類問題討論的限制，並以一組模擬數據與眞實試驗資料來加以闡述這程式之應用，目的是讓使用者能簡便地計算出種子傳毒率之數值，以利流行病學研究評估其經濟臨界值，並據以研擬病毒種子傳毒比率之最高容許度，最後建立種子健康檢查標準流程及與國際接軌的驗證體系，以落實防疫檢疫政策之執行。

關鍵詞：容許度、取樣、群體測試、種子傳毒率、適合度檢定

建立種子健康檢查的標準流程 (SOP)，除了檢測技術的開拓外，對於病毒種傳率之最高容許度 (tolerance level) 必須充分瞭解，才能與國際的植物健康驗證 (phytosanitary certification) 制度接軌，而種傳率估計值是執行種傳忍受極限之重要依據，因此本文主要是探討並發展一種統計分析的方法來估算群體測試中的種傳率。

首先，在取樣方式上種子病害檢測與一般種子檢查是相同的，首重取樣均勻，且能代表整批種子[2, 10]。因此遵循國際種子檢查規則的取樣程序，由同一批號的種子批 (seed lot) 從不同位置逢機抽取原始樣品 (primary sample) ，其取樣頻度 (sampling intensity) 依規則所定。再將所有原始樣品組合並混合成複合樣品 (composite sample)，由部份或全部複合樣品組成報驗樣品 (submitted sample) 送往檢測機關。檢測機關實驗室從報驗樣品中抽取出來進行檢測的次樣品 (sub-sample) 即爲供試樣品 (working sample)[1, 7]。

在計算種傳率時，先將待測之供試樣品種子數依 n 值加以區隔成 k 個群 (即在每個群內有 n 粒種子)，每一群內所有之種子樣品集合共同萃取成單一樣本進行檢測，若偵測到病毒陽性反應之群聚數爲 r，則該批種子之種傳率 θ 即可表示爲如下公式：

$$\theta = 1 - e^{\frac{\ln(1-p)}{n}} \qquad (1)$$

其中 $p = \frac{r}{k}$ 代表所得結果中偵測到病毒陽性反應之比率[5, 6]。

當不同種子數的樣本被檢驗時 (即 n 爲不同) ，對於每個樣本可能得到不同的估計值，對於第 i 個樣本偵測出陽性反應的機率 (Pi) 與種傳率 (θ)，之間的關係如下：

$$\ln[-\ln(1-p_i)] = \ln(n_i) + \ln[-\ln(1-\theta)] \qquad (2)$$

此爲 complementary log-log (c-log-log) 連接函數 (link function)，ln (-1n (1- θ)) 爲截距項，解釋變數爲 1n (n_i)，但其係數必須爲1，據此估算出種傳率 (θ) 之點估計值、第 i 個樣本之陽性反應配適值 (fitted value)、適合度檢定 (goodness of fit) 並以最大概度估計法 (maximum likelihood estimation) 之大樣本變異數來建構種傳率的 95% 信賴區間[9]，上述結果皆可由英國園藝研究機構 (Horticulture Research International) 在 1995 年所開發之種子試驗分析程式 (seed test analysis program, *STpro*) 來執行[8]。但此程式有一個嚴重問題，即當資料經過適合度檢定，若通過檢定即表示用 c-log-log 連接函數配適相當不錯；若未通過檢定即配適不好時 (p-value 相當小)，此程式仍繼續使用 c-log-log 連接

函數，而非嘗試不同的連接函數，因此本文即針對此一問題，使用 R 程式語言開發出更具有彈性的連接函數，以便順利解決此一問題 (程式名稱爲“seedhealth”)。

在統計文獻中，除了 c-log-log 之外，probit 與 logit 連接函數皆可配適二元 (binary) 資料，在此探討之例子爲健康抑或帶毒種子；一般而言，c-log-log 連接函數是不對稱的，p 在接近 0 是較和緩，但接近 1 較急速；probit 常用於毒理學之試驗，而 logit 在醫學上則廣泛被運用，後兩者連接函數是對稱的，亦即接近 0 與 1 之速度類似[3]。

在本文所發展的程式之流程如下：

(1) 先輸入適合度檢定所允許最大之型一誤差 (α)，即原始假說 Ho 是對的卻被拒絕掉所犯之錯誤機率，執行 c-log-log 連接函數之配適，若 p-value > α，則程式列印種傳率之相關數值；但若 p-value ≤ α，則拒絕 Ho，表示 c-log-log 連接函數不適當，則進行下一步驟。

(2) 接續執行 probit 連接函數配適，若 p-value > α，則程式列印種傳率之相關數值；但若 p-value ≤ α，則拒絕 Ho，表示 probit 連接函數不適當，則進行下一步驟。

(3) 接續執行 logit 連接函數配適，若 p-value > α，則程式列印種傳率之相關數值。

近期發展出的 R 軟體是一個免費的統計分析軟體，其最大的優點爲取得軟體的便利性，並具有一般統計軟體 (例如：SAS) 分析數據的強大功能，因此最近幾年在醫學領域上的運用相當普遍。R 有各種版本，可以在 UNIX 版本、MacOS X 版本和其他 Windows 版本 (95 and later) 的作業系統上運行。R 軟體的主要介面是互動交談命令模式，使用者必須先熟悉 R 的程式語言語法，才能執行 R 的各項功能。它是一套擁有許多機率統計功能的程式語言。並非直接點選就可以操作的統計軟體。以下分別陳述 R 軟體安裝步驟及我們所寫的程式 (seedhealth.txt) 在其介面執行之方法：

安裝步驟:

Step 1: 下載 R 程式

http://cran.us.r-project.org/bin/windows/base/ or http://cran.csie.ntu.edu.tw/bin/windows/base/ (下載 R - 2.5.1-win32.exe)

Step 2: 執行R-2.5.1-win32.exe 安裝

安裝完成後, 執行桌面上的捷徑 R

Step 3:

將程式 seedhealth.txt 檔複製並貼上 R 軟體內，內有文章中所使用的數據可供測試。

Step 4:

茲就本文模擬之首例所輸入的資料格式加以解釋，數據輸入方式如下：

seedhealth (c( ), c( ), c( ), alpha)

在 c( ) 內依照順序填入，中間需以「,」間隔，第一個爲每個試驗的樣本大小數目，第二個爲每個試驗執行的次數，第三個爲每個試驗執行後測得 positive 次數，alpha 爲適合度檢定所能容許之型一誤差，可視情況來訂定，一般常以 0.05 定之。

以一組模擬之資料來驗證此程式之結果 (表一)，群體檢測一批種子有4種不同之種子數，分別爲 n = 50, 100, 250 與 500 在一個群體樣本中，在檢測 k = 5 群 n = 50 粒種子中其中有 r = 1 群呈現陽性反應，在檢測 10 群 100 粒種子中其中有 5 群呈現陽性反應，在檢測 15 群 250 粒種子中其中有 13 群呈現陽性反應，在檢測 20 群 500 粒種子中其中有 15 群呈現陽性反應，根據此資料，在 R 的界面上利用 seedhealth 程式計算出種傳率如下：> seedhealth (c(50,100,250,500),c(5,10,15,20), c(1,5,13,15),0.05)

```
****c-log-log link function is suitable for the dataset. ***
******Its P-value of goodness of fit = 0.0707 *******
Estimated proportion of infected seeds 0.0044
The upper 95% confidence limits 0.0061
The lower 95% confidence limits 0.0027
```

	No. seeds in sample	No. samples tested	No. positive	Fitted value
[1,]	50	5	1	0.9824
[2,]	100	10	5	3.5436
[3,]	250	15	13	9.9758
[4,]	500	20	15	17.7562

當 α 設定爲 0.05 時，c-log-log 連接函數配適結果之適合度檢定 p-value = 0.0707，因 0.0707 > 0.05，故程式就此停住，其種傳率爲 0.0044，95% 信賴區間爲 (0.0027, 0.0061)，4 種不同種子數的樣本呈現陽性反應之預測依序爲 0.9824、3.5436、9.9758 和 17.7562。若 α 設定爲 0.08 時，c-log-log 連接函數配適結果之適合度檢定 p-value = 0.0707 < 0.08，故程式尋求 probit 連接函數來配適，結果其 p-value = 0.0995 > 0.08，probit 配適結果不錯，其種傳率爲 0.0073，95% 信賴區間爲 (0.0043, 0.0102)，4 種不同種子數的樣本呈現陽性反應之預測依序爲 0.7817、3.7581、10.8851 和 18.0397。若 α 設定爲 0.10 時，c-log-log 與 probit 連接函數之適合度檢定皆小於 0.10，故程式尋求 logit 連接函數來配適，結果其 p-value = 0.3463 > 0.10，logit 配適結果不

錯，其種傳率為 0.0099，95% 信賴區間為 (0.0036, 0.0162) ，4 種不同種子數的樣本呈現陽性反應之預測依序為1.6611、4.9874、10.6989 和 16.6527 (表一)。

另有實例以 ELISA 群體試驗檢測種子帶毒率：玉米種子一批，所有送檢樣本的種子經浸水軟化並在濕室中催芽 3 天，從中取 135 株已發芽的種子，分別剝取種皮為一組，和含胚體的殘餘組織樣本為另一組，集 3 粒種子為一個檢測單位，進行間接法 ELISA。檢驗用抗體為自製玉米矮化嵌紋病毒 (*Maize dwarf mosaic virus*, MDMV) 抗血清 [4]。出芽試驗 : 催芽後的種子分別種植於 72 格穴盤中，8 天後逐株觀察病徵並剪取葉片進行間接法ELISA，檢測種子傳播 MDMV 至幼苗的情形。結果在 45 粒種皮檢測單位中有 1 粒為陽性反應 (有帶 MDMV) ，但含胚體組織全為陰性反應 (未帶 MDMV) 。出芽試驗結果，在 135 株苗中確有 1 株為陽性反應 (已感染 MDMV)，且可看到其病徵。綜合兩種試驗的結果，經以 seedhealth 程式計算，種傳百分比率估計為 0.74%，其 95% 信心範圍為 0% - 1.77%，對於 c-log-log 連接函數之適合度 chi-squared statistic = 0.00 (1 df; P = 0.9958) (表二)。

本文擴充英國園藝研究機構所開發之種子試驗分析程式 *STpro*，以不同之連接函數 (c-log-log, probit 與 logit) 較有彈性來配適資料，目的在尋求最佳之結果。但在計算種傳率之問題上，仍有些問題急待解決，分

表一、模擬研究之資料 (A) 與執行程式之結果 (B)

Table 1. The simulated data (A) and the result by implementing the code ("seedhealth") (B)

(A)

No. of seeds in sample	No. of samples tested	No. of positive
50	5	1
100	10	5
250	15	13
500	20	15

(B)

	Link function	P-value of Goodness of fit	Seed Transmission rate	95% Confidence interval	Fitted value
α= 0.05*	c-log-log	0.0707	0.0044	(0.0027, 0.0061)	0.9824
					3.5436
					9.9758
					17.7562
α= 0.08*	probit	0.0995	0.0073	(0.0043, 0.0102)	0.7817
					3.7581
					10.8851
					18.0397
α= 0.10*	logit	0.3463	0.0099	(0.0036, 0.0162)	1.6611
					4.9874
					10.6989
					16.6527

* αrepresents making the probability of a type I error for goodness of fit test.

表二、玉米種子 MDMV 檢測結果與種傳率之估計

Table 2. Result of testing MDMV and estimation of seed transmission rate for corn seeds

Testing method	No. of seeds in sample	No. of samples tested	No. of positive	Fitted value
ELISA group test	3	45	1	0.9963
Growing-on test	1	135	1	1.0037

Estimated proportion of infected seeds = 0. 74%

95 % confidence limits 0 % to 1.77%

Goodness-of-fit chi-squared statistic for c-log-log = 0.00 (1 df; P = 0.9958).

別說明如下：

(1) 假設在數個不同種子數之樣本其種傳率 (θ) 是固定的，但在實際情形有時 θ 非固定值，例如：當使用不同之種子檢測技術時，θ 可能會因不同方法而有所不同，這種情形應修正公式 (2) 而使用不同之種傳率 (θ_i)，且據此可看出不同之檢測技術方法是否有所差異。

(2) 一般而言，種子試驗資料分析之理論基礎爲從一批試驗之種子採用隨機抽樣 (random sampling)，每粒種子有相同的機會被分類爲健康或帶毒的 (本文之理論基礎亦此)，但事實上種子帶毒率可能有群聚 (clustering) 現象。

(3) 經試驗證實罹病植株的病毒之種子傳播率一般偏低 (<1%) [2]，況且對未知的一整批種子，在一次檢驗中最少須要檢測幾個樣本？將是未來研究採樣技術之課題。

(4) 在種子檢測中，常使用群體測試法 (group testing)，但是因爲種子內所帶的病毒數量少且比率低，所以常有僞陰性的產生，這種現象可能是混合種子數量 (n) 進行檢測的個數不夠，抑或檢測之群數 (k) 不足，導致無法測出病毒，所以如何求得最佳之 n 與 k，亦將是另一個重要的研究方向。

具有經種子傳播能力的病毒在農業生態和檢防疫上有其特殊性，而種子帶毒的診斷與防除技術之重要性已不言可喻。上述實例中MDMV 雖非感染至胚的種傳模式，但是 MDMV 確實存在種子中 (seed-borne)，且可傳染至下一代苗株上。雖然 MDMV 未被 Albrechtsen 列入重要種傳病毒 [2]，但是在玉米產區的病害管理，MDMV 的種傳現象卻是不能忽視的一個重要因素。

引用文獻 (LITERATURE CITED)

1. 鄧汀欽、蔡錦慧、廖吉彥. 2005. 瓜類種子傳播病毒病害的特性. 植物種苗 7:1-19。
2. Albrechtsen, S. E. 2006. Testing Methods For Seed-transmitted Viruses: Principles and Protocols. CABI Publishing, Wallingford, UK. 288 pp.
3. Collett, D. 2003. Modelling binary data. 2nd ed. Chapman & Hall/CRC, 387 pp.
4. Deng, T. C., and Huang, C. H. 1986. Purification and antiserum preparation of maize dwarf mosaic virus. J. Agri. Res. China 35:350-359.
5. Maury, Y., Bossennec, J. M., Boudazin, G., Hampton, R. O., Pietersen, G., and Maguire, J. D. 1987. Factors influence ELISA evaluation of transmission of pea seed borne mosaic virus infected seeds. Agronomie 7:225-230.
6. Maury, Y., Duby, C., and Khetarpal, R. K. 1998. Seed certification for viruses. In: A. Hadidi, et al., eds. Plant Virus Disease Control. APS Press, St. Paul, Minnesota, USA., pp. 237-248.
7. Morrison, R. H. 1999. Sampling in seed health testing. Phytopathology 89:1084-1087.
8. Ridout, M. S., and Roberts, S. J. 1995. Instructions and notes on *STpro* program. Horticulture Research International, UK. 8 pp. (Website: www.planthealth.co.uk\downloads)
9. Roberts, S. J., Phelps, K., Taylor, J.D. and 1993. Design and interpretation of seed health assays. In: J. W. Sheppard, ed., Proceedings of the First ISTA Plant Disease Committee Symposium on Seed Health Testing, Ottawa, Canada, August 1993., pp.115-125.
10. Russell, T. S. 1988. Some aspects of sampling and statistics in seed health testing and the establishment of threshold levels. Phytopathology 78:880-881.

ABSTRACT

Chiang, K. S.[1], Deng, T.-C.[2,3], Lai, H. H.[1], and Wang, S. F.[1] A statistical method to estimate virus seed-transmission rate by results of group tests. Plant Pathol. Bull. 16: 225-229 ([1]Department of Agronomy, National Chung Hsing University; [2]Plant Pathology Division, Agricultural Research Institute, COA, Executive Yuan; [3]Corresponding author, Email: tcde@wufeng.tari.gov.tw; Fax: 04-2333-8162)

To develop the economical and useful techniques for detecting seed-transmitted viruses and to estimate the seed transmission rates of seed lots by statistical methods, we develop a code to calculate the seed transmission rate flexibly by the result of group test. The code can be utilized to solve the problem and limitation for the discussion of the previous literatures involved this topic. Besides, a simulated dataset and a real dataset will be illustrated. The code will be capable of serving as a research tool to assess the economic threshold, set up the maximal tolerance level of the seed transmission rate, and standardize the procedure in testing protocols for establishing a worldwide phytosanitary system. In addition, it will help the quarantine authority to execute the policy for avoiding the phytosanitary risks.

Key words: tolerance level, sampling, group test, seed transmission rate, goodness of fit